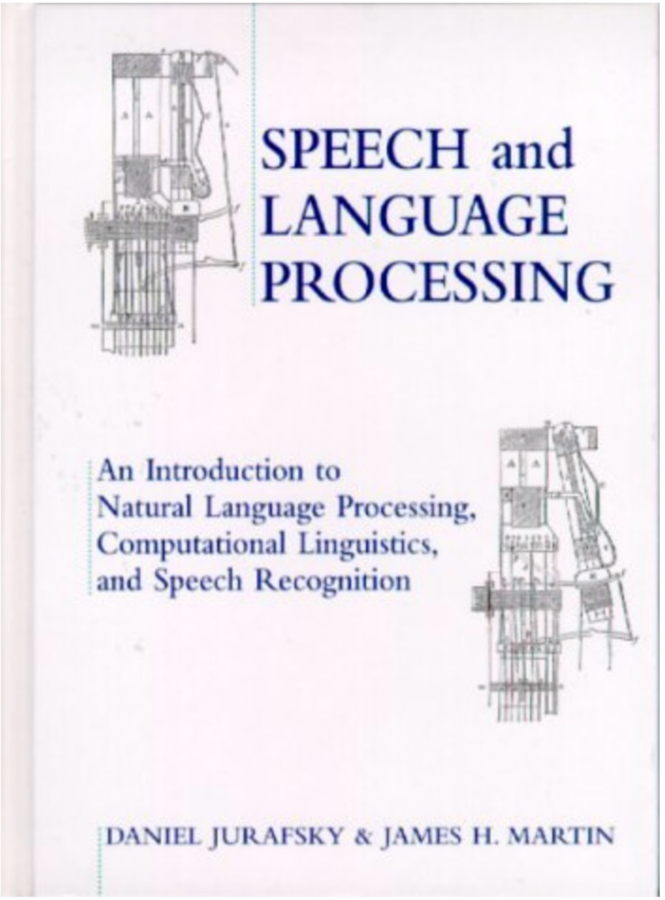
The Parable of the Parser

Ross Girshick

AI2

CV 20/20 Retrospective Vision Workshop, CVPR 2024

NLP circa 1999



1st edition

Summary of Contents

1 Introduction 1

I Words 19

2 Regular Expressions and Automata 21
3 Morphology and Finite-State Transducers 57
4 Computational Phonology and Text-to-Speech 91
5 Probabilistic Models of Pronunciation and Spelling 139
6 $\boxed{\text{N-grams}}$ 189
7 HMMs and Speech Recognition 233

II Syntax 283

8 Word Classes and Part-of-Speech Tagging 285
9 Context-Free Grammars for English 319
10 $\boxed{\text{Parsing with Context-Free Grammars}}$ 353
11 Features and Unification 391
12 Lexicalized and Probabilistic Parsing 443
13 Language and Complexity 473

III Semantics 495

14 $\boxed{\text{Representing Meaning}}$ 497
15 Semantic Analysis 543
16 Lexical Semantics 587
17 Word Sense Disambiguation and Information Retrieval 627

IV Pragmatics 661

18 Discourse 663
19 Dialogue and Conversational Agents 715
20 Generation 759
21 Machine Translation 797
A Regular Expression Operators 829
B The Porter Stemming Algorithm 831
C C5 and C7 tagsets 835
D Training HMMs: The Forward-Backward Algorithm 841

Bibliography 851

Index 923

words → syntax (parsing) → semantics → pragmatics

Comp. linguistics

AI

NLP today has transform(er)ed

Speech and Language Processing

An Introduction to Natural Language Processing,
Computational Linguistics, and Speech Recognition

Third Edition draft

Daniel Jurafsky
Stanford University

James H. Martin
University of Colorado at Boulder

Draft of February 3, 2024. Comments and typos welcome!

3rd edition

Summary of Contents

I Fundamental Algorithms for NLP 1
1 Introduction 3
2 Regular Expressions, Text Normalization, Edit Distance 4
3 N-gram Language Models 32
4 Naive Bayes, Text Classification, and Sentiment 60
5 Logistic Regression 81
6 Vector Semantics and Embeddings 105
7 Neural Networks and Neural Language Models 136
8 Sequence Labeling for Parts of Speech and Named Entities 162
9 RNNs and LSTMs 187
10 Transformers and Large Language Models 213
11 Fine-Tuning and Masked Language Models 242
12 Prompting, In-Context Learning, and Instruct Tuning 263
II NLP Applications 265
13 Machine Translation 267
14 Question Answering and Information Retrieval 293
15 Chatbots & Dialogue Systems 315
16 Automatic Speech Recognition and Text-to-Speech 337
III Annotating Linguistic Structure 365
17 Context-Free Grammars and Constituency Parsing 367
18 Dependency Parsing 391
19 Information Extraction: Relations, Events, and Time 415
20 Semantic Role Labeling 441
21 Lexicons for Sentiment, Affect, and Connotation 461
22 Coreference Resolution and Entity Linking 481
23 Discourse Coherence 511
Bibliography 533
Subject Index 563

The book is now 40% shorter

Neural networks, LMs, transformers, and LLMs

Misc. other stuff (including parsing)

1999 → 2024: What happened to parsing?

“*In the early history of NLP these structures [parses] were an intermediate step toward deeper language processing.*

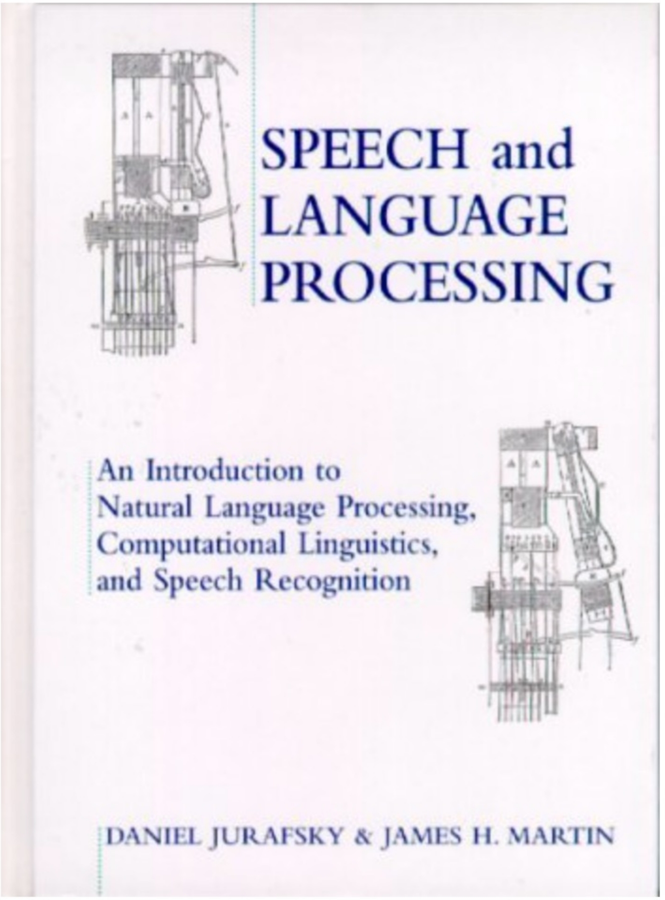
1999 → 2024: What happened to parsing?

"*In the early history of NLP these structures [parses] were an intermediate step toward deeper language processing.*

In modern NLP, we don't generally make explicit use of parse or other structures inside the neural language models [...], or directly in applications like those we discussed in Part II.

- Jurafsky & Martin

NLP circa 1999



1st edition

Summary of Contents

1 Introduction 1
I Words 19
2 Regular Expressions and Automata 21
3 Morphology and Finite-State Transducers 57
4 Computational Phonology and Text-to-Speech 91
5 Probabilistic Models of Pronunciation and Spelling 139
6 N-grams 189
7 HMMs and Speech Recognition 233
II Syntax 283
8 Word Classes and Part-of-Speech Tagging 285
9 Context-Free Grammars for English 319
10 Parsing with Context-Free Grammars 353
11 Features and Unification 391
12 Lexicalized and Probabilistic Parsing 443
13 Language and Complexity 473
III Semantics 495
14 Representing Meaning 497
15 Semantic Analysis 543
16 Lexical Semantics 587
17 Word Sense Disambiguation and Information Retrieval 627
IV Pragmatics 661
18 Discourse 663
19 Dialogue and Conversational Agents 715
20 Generation 759
21 Machine Translation 797
A Regular Expression Operators 829
B The Porter Stemming Algorithm 831
C C5 and C7 tagsets 835
D Training HMMs: The Forward-Backward Algorithm 841
Bibliography 851
Index 923

(sub)words

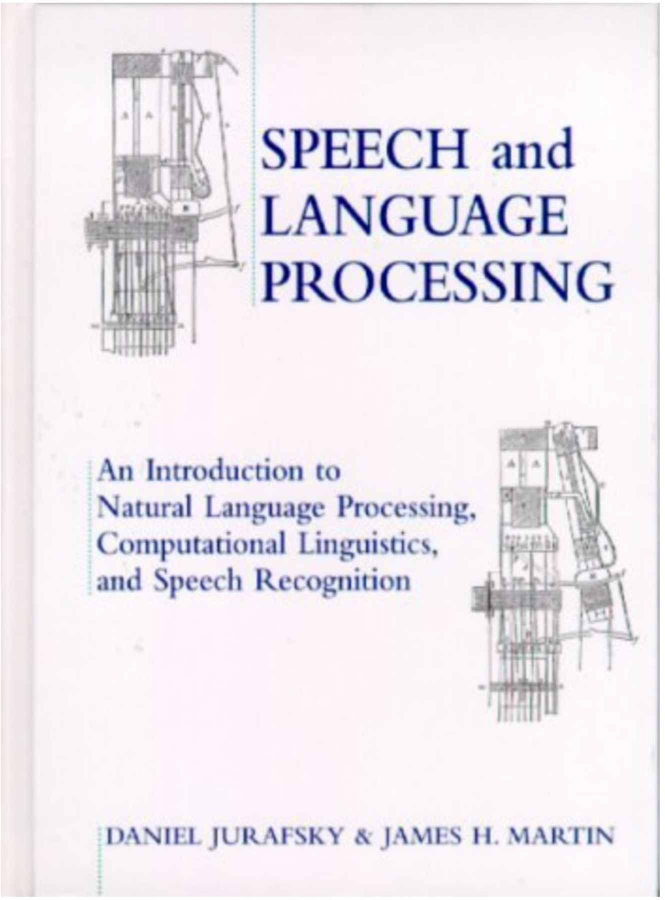
syntax (parsing)

semantics

pragmatics

LLMs

NLP circa 1999



1st edition

Summary of Contents

1 Introduction 1
I Words 19
2 Regular Expressions and Automata 21
3 Morphology and Finite-State Transducers 57
4 Computational Phonology and Text-to-Speech 91
5 Probabilistic Models of Pronunciation and Spelling 139
6 N-grams 189
7 HMMs and Speech Recognition 233
II Syntax 283
8 Word Classes and Part-of-Speech Tagging 285
9 Context-Free Grammars for English 319
10 Parsing with Context-Free Grammars 353
11 Features and Unification 391
12 Lexicalized and Probabilistic Parsing 443
13 Language and Complexity 473
III Semantics 495
14 Representing Meaning 497
15 Semantic Analysis 543
16 Lexical Semantics 587
17 Word Sense Disambiguation and Information Retrieval 627
IV Pragmatics 661
18 Discourse 663
19 Dialogue and Conversational Agents 715
20 Generation 759
21 Machine Translation 797
A Regular Expression Operators 829
B The Porter Stemming Algorithm .... [illegible]
C C5 and C7 tagsets .... [illegible]
D Training HMMs: The Forward-Backward Algorithm .... [illegible]
Bibliography [illegible]
Index 923

Fake Tasks

Real Tasks

(sub)words

syntax (parsing)

semantics

pragmatics

LLMs

If your end goals are chat, MT, etc., then do not build a parser.

(Of course, if your end goal is to have a parser, then build a parser. But likely you will do this with an LLM anyway.)

What are the “parsers” of computer vision?

- Before we answer this question, take a step back
- We need to understand what our real tasks are
- Let’s study the rise of LLMs
  - What lessons can we learn?

Why are LLMs so successful?

- *Real* tasks are text generation tasks
  - High societal and economic value
  - Universality: can do POS tagging, parsing, etc. *as text generation* (if you *really* want to)

NLP circa 2024

Summary of Contents

1 Introduction ... 1
I Words 19
2 Regular Expressions and Automata ... 21
3 Morphology and Finite-State Transducers ... 57
4 Computational Phonology and Text-to-Speech ... 91
5 Probabilistic Models of Pronunciation and Spelling ... 139
6 N-grams ... 189
7 HMMs and Speech Recognition ... 233
II Syntax 283
8 Word Classes and Part-of-Speech Tagging ... 285
9 Context-Free Grammars for English ... 319
10 Parsing with Context-Free Grammars ... 353
11 Features and Unification ... 391
12 Lexicalized and Probabilistic Parsing ... 443
13 Language and Complexity ... 473
III Semantics 495
14 Representing Meaning ... 497
15 Semantic Analysis ... 543
16 Lexical Semantics ... 587
17 Word Sense Disambiguation and Information Retrieval ... 627
IV Pragmatics 661
18 Discourse ... 663
19 Dialogue and Conversational Agents ... 715
20 Generation ... 759
21 Machine Translation ... 797
A Regular Expression Operators ... 829
B The Porter Stemming Algorithm ... 831
C C5 and C7 tagsets ... 835
D Training HMMs: The Forward-Backward Algorithm ... 841
Bibliography 851
Index 923

Fake Tasks

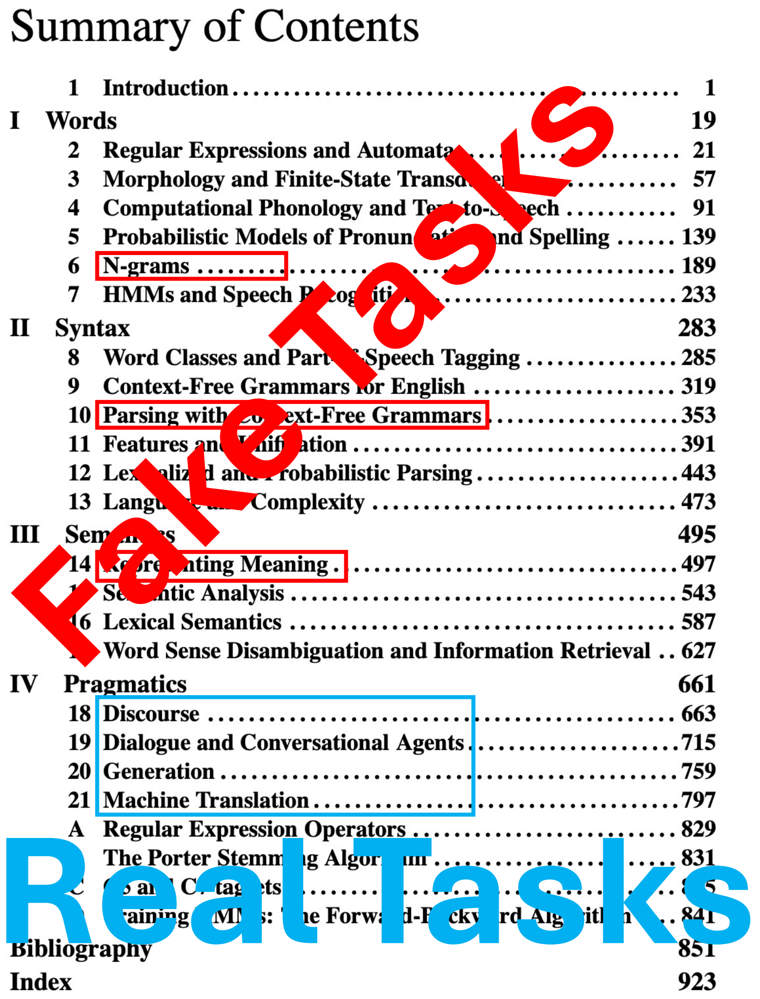
Real Tasks

NLP circa 1999

Why are LLMs so successful?

- *Real* tasks are text generation tasks
  - High societal and economic value
  - Universality: can do POS tagging, parsing, etc. *as text generation* (if you *really* want to)
- **Text generation is all you need**

NLP circa 2024



Summary of Contents

	1	Introduction	1
I		Words	19
	2	Regular Expressions and Automata	21
	3	Morphology and Finite-State Transducers	57
	4	Computational Phonology and Text-to-Speech	91
	5	Probabilistic Models of Pronunciation and Spelling	139
	6	N-grams	189
	7	HMMs and Speech Recognition	233
II		Syntax	283
	8	Word Classes and Part-of-Speech Tagging	285
	9	Context-Free Grammars for English	319
	10	Parsing with Context-Free Grammars	353
	11	Features and Unification	391
	12	Lexicalized and Probabilistic Parsing	443
	13	Language and Complexity	473
III		Semantics	495
	14	Representing Meaning	497
	15	Semantic Analysis	543
	16	Lexical Semantics	587
	17	Word Sense Disambiguation and Information Retrieval	627
IV		Pragmatics	661
	18	Discourse	663
	19	Dialogue and Conversational Agents	715
	20	Generation	759
	21	Machine Translation	797
	A	Regular Expression Operators	829
	B	The Porter Stemming Algorithm	831
	C	C5 and C7 tagsets	835
	D	Training HMMs: The Forward-Backward Algorithm	841
		Bibliography	851
		Index	923

NLP circa 1999

Why are LLMs so successful?

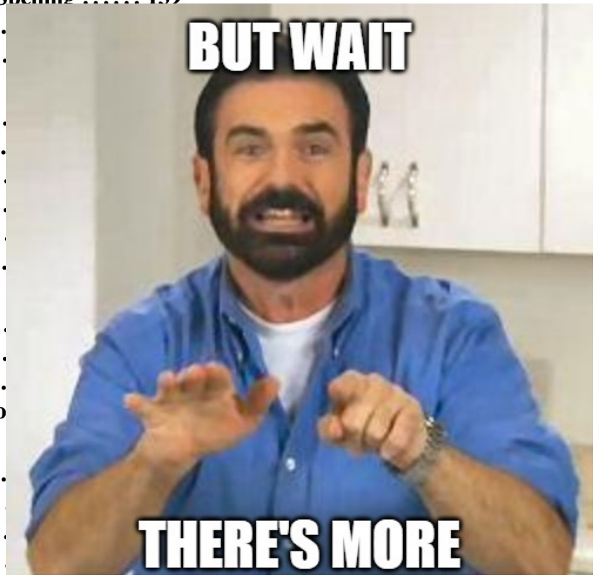
- *Real* tasks are text generation tasks
    - High societal and economic value
    - Universality: can do POS tagging, parsing, etc. *as text generation* (if you *really* want to)
- **Text generation is all you need**
- **But wait, there's more!**
  What if I told you:
    - *<u>High-quality data is abundant</u>*
    - *<u>The test task = the training task</u>*
    - *<u>Training is basically supervised</u>*

NLP circa 2024

Summary of Contents

1 Introduction ... 1
I Words 19
2 Regular Expressions and Automata ... 21
3 Morphology and Finite-State Transducers ... 57
4 Computational Phonology and Text-to-Speech ... 91
5 Probabilistic Models of Pronunciation and Spelling ... 139
6 N-grams ...
7 HMMs and Speech Recognition ...
II Syntax
8 Word Classes and Part-of-Speech Tagging ...
9 Context-Free Grammars for English ...
10 Parsing with Context-Free Grammars ...
11 Features and Unification ...
12 Lexicalized and Probabilistic Parsing ...
13 Language and Complexity ...
III Semantics
14 Representing Meaning ...
15 Semantic Analysis ...
16 Lexical Semantics ...
17 Word Sense Disambiguation and Informatio
IV Pragmatics
18 Discourse ...
19 Dialogue and Conversational Agents ...
20 Generation ...
21 Machine Translation ...
A Regular Expression Operators ... 829
B The Porter Stemming Algorithm ... 831
C C5 and C7 tagsets ... 835
D Training HMMs: The Forward-Backward Algorithm ... 841
Bibliography 851
Index 923

Fake Tasks



Real Tasks

NLP circa 1999

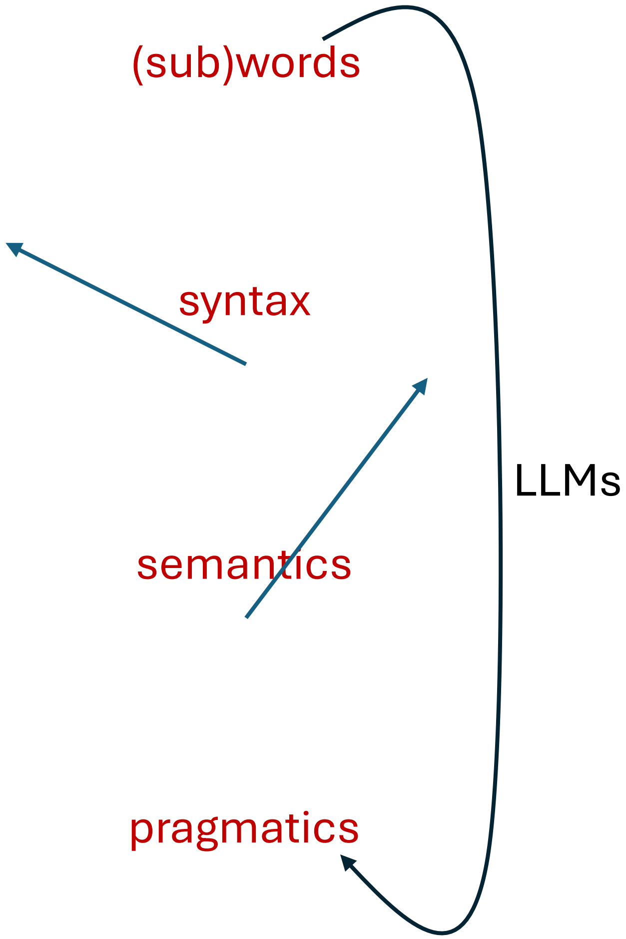
The LLM miracle

- Test task = training task (... maybe also data? 🤔)
- Massive high-quality data
- Supervised training
- All for what people truly care about: *text generation*

🤯

- This basically never happens, and it should blow our minds
  - Most CV/ML: *sad proxy loss/tasks, small data, low annotation quality, ...*

The LLM miracle



- Test task = training task (... maybe also data? 🤔)
- Massive high-quality data
- Supervised training
- All for what people truly care about: *text generation*

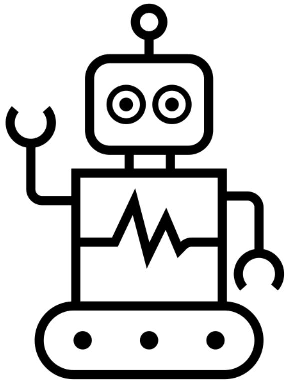
🤯

- This basically never happens, and it should blow our minds
  - Most CV/ML: *sad proxy loss/tasks, small data, low annotation quality, ...*

What are the real tasks of computer vision?

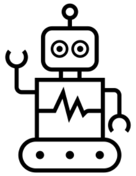
- Caveat: focusing on recognition-y things, not 3D-y things

- Q1: Who benefits from CV?
- A1: Those who cannot see
  - People with low or no vision
    - Real task = open-ended visual QA about the real world
  - In general, AI agents operating in visual environments
    - Real tasks =
      - Robots: “cook me dinner”, “fold my laundry”, “wash my dishes”, ...
      - Internet agents: “shop for my climbing shoes at a good sale price”, ...
      - ...



What are the real tasks of computer vision?

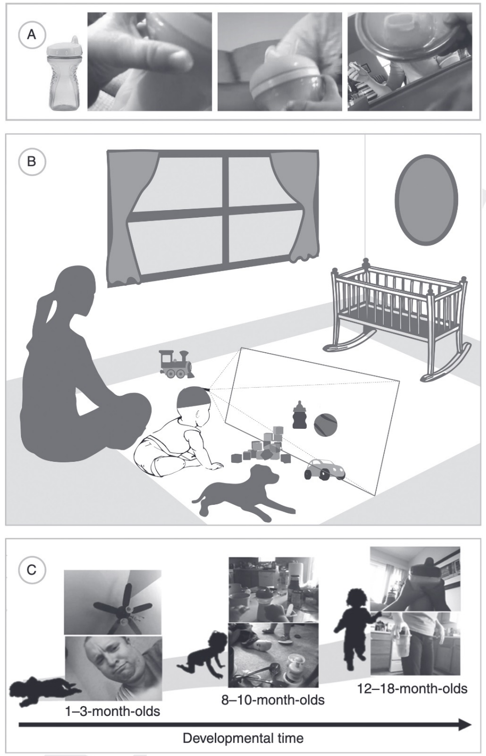
- Real tasks =
  - Open-ended visual QA about the real world
  - Robot: “cook me dinner”, …
  - Internet agent: “shop for me”, …



- Traditional CV tasks (classification, detection, segmentation, …) are not required *a priori*
  - Helpful intermediates or *The Wrong Way* (i.e., parsers)?

What are the real tasks of computer vision?

- Q2: What is the most scientifically important direction?
- A2: Algorithms that learn with human-like data constraints
  - Constraints
    - Ego-centric video
    - ~24M frames in 18 months (12 hours / day, 1 FPS)
    - Limited embodied control
    - Observations have very different statistics cf. web data
  - Long term, this is probably way more important
  - It doesn’t require a data miracle
  - But it’s hard, we have no gradient, few people work on it, and maybe not needed (birds *vs.* airplanes) …



Jayaraman and Smith

What are the “parsers” of computer vision?

- Identify *your* real tasks (I gave 2)
  - General VQA and embodied vision
  - Learning with human-like data constraints
- Given your real tasks, fake tasks (“parsers”) are those subproblems that are not helpful for solving your real tasks
  - This definition is relative to your choice of real tasks

Let’s take general VQA and embodied vision

- The system needs a *vast* array of skills
  - Scene text reading (OCR)
  - Object recognition (classification)
  - Object delineation (detection, segmentation)
  - Chart / infographic parsing
  - Document parsing
  - Instrument reading
  - Place recognition
  - Action recognition
  - Face recognition
  - World knowledge
  - …

Your Real Tasks

Let's take general VQA and embodied vision

Your Fake Tasks

- The system needs a *vast* array of skills
  - Scene text reading (OCR)
  - Object recognition (classification)
  - Object delineation (detection, segmentation)
  - Chart / infographic parsing
  - Document parsing
  - Instrument reading
  - Place recognition
  - Action recognition
  - Face recognition
  - World knowledge
  - …

Bespoke solutions to these subproblems will never fit together to solve general VQA and embodied vision

What’s the “LLM miracle” story here?

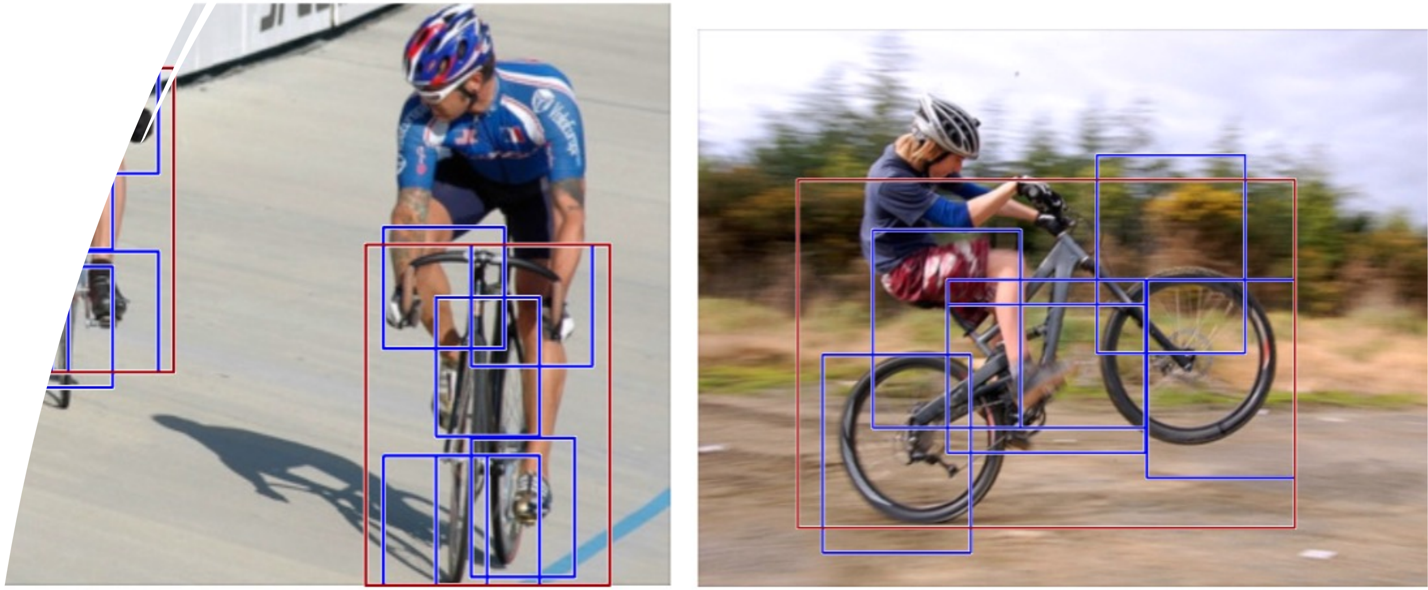
- Hypothesis: scaling vision backbones plus LLMs (simple LLaVA-style model) will effectively “solve” the general VQA / embodied vision tasks to the extent that LLMs “solve” any problem

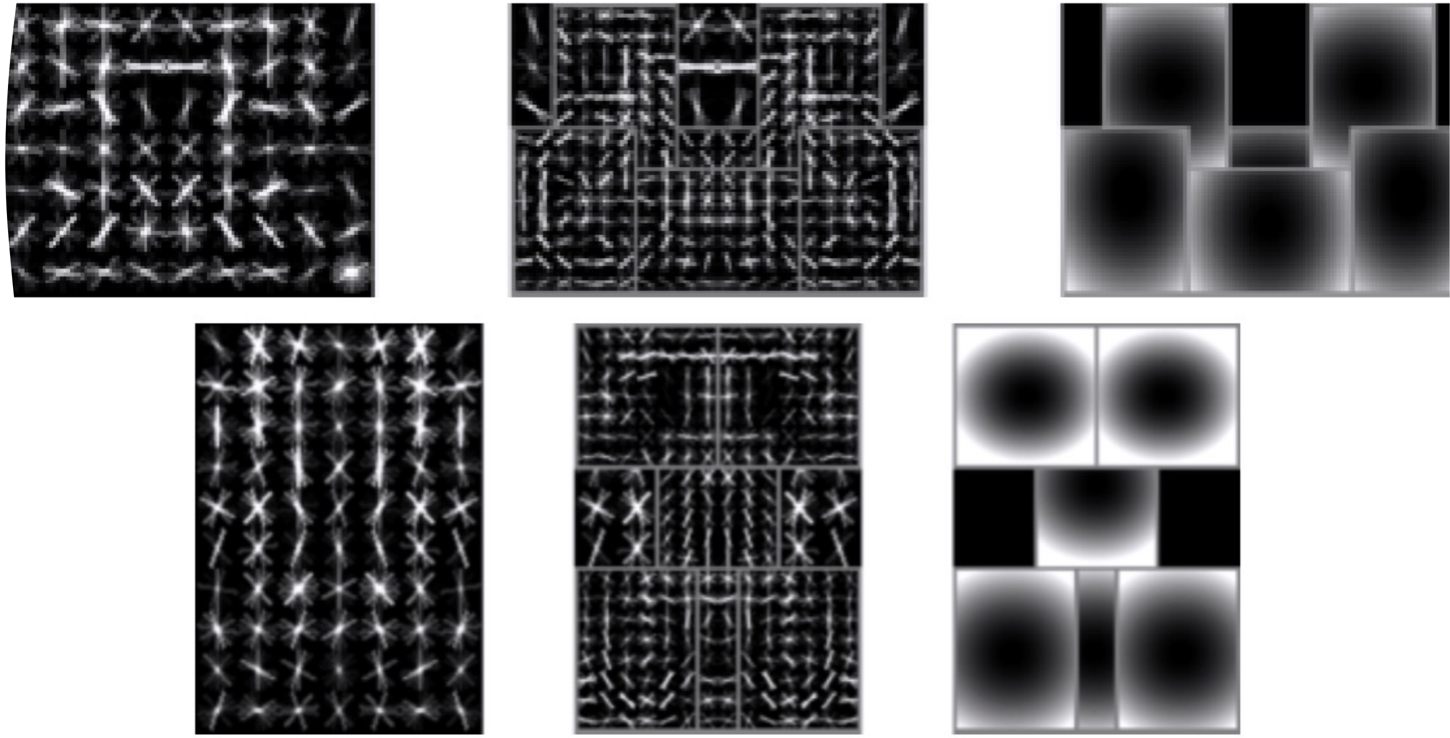
- But, the big open question is what data to use and how to get it

- It’s not as clean & easy as for LLMs, but I think it’s possible

Reflections on detectors as parsers

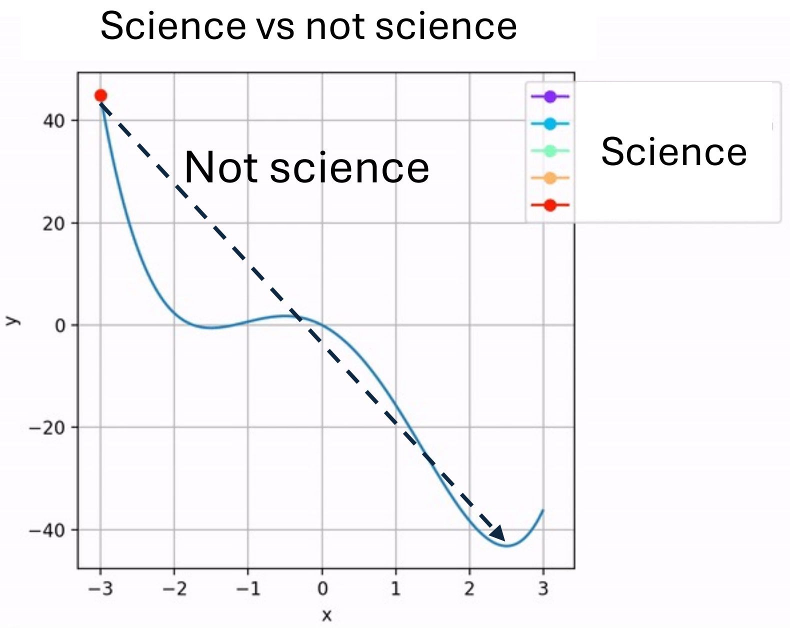
- I started working on object detection in 2008 and worked on it until ~2022
- The "why?" didn't matter for a long time
- *Detection didn't work at all*, it was interesting to make anything work at all





Was working on detection a mistake?

- Absolutely not
  - Science is iterative and noisy
  - A single step to the optimum is magic
- We gained knowledge along the way
  - Our modern techniques build on the knowledge we gained
- But to advance, we must iterate
  - Datasets must evolve
  - *Tasks must evolve*



Takeaway 1

- Be skeptical of ideas taken for granted for the last 50-60 years
- Example: object detection
  - I want to solve open-ended, real-world QA that powers embodied agents
  - Making better object detectors is 100% the wrong direction for this achieving this goal, imo
  - It’s too limited, too brittle, too data constrained, etc.

Takeaway 2

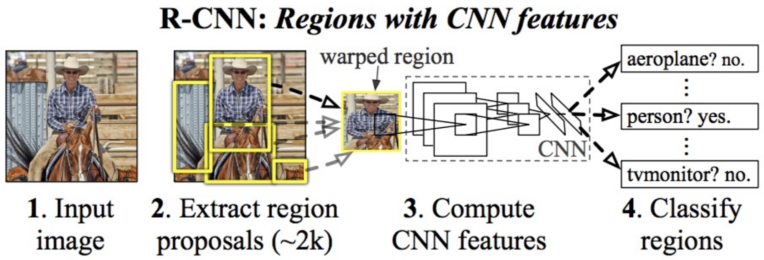
- Answer “what are the real tasks?” for yourself
  - We need diversity of perspectives, most will be wrong
  - It’s very dull when everyone thinks in the same way
  - *Your* fake tasks are defined relative to *your* real tasks

- Examples:
  - General VQA and embodied vision
  - Learning with human-like data constraints
  - …
  - *What are yours?*

Takeaway 3

- Trying to solve a scientifically interesting problem with no other motivation than gaining knowledge can be extremely fruitful

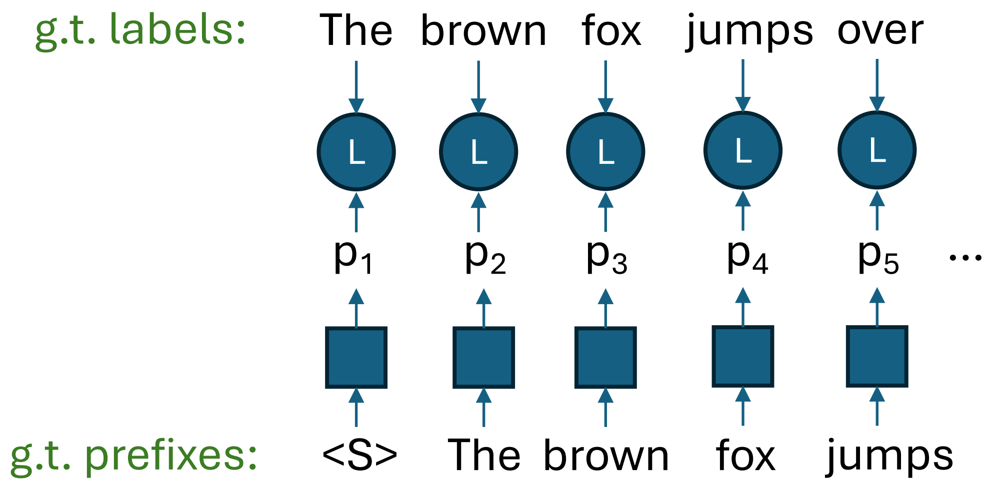
- Example: object detection
  - Key early success of deep learning in CV, after classification
  - Moved the CV community into deep learning, convinced many people



R-CNN paper from CVPR 2014

Thank you!

Data curation makes this supervised learning



- In **self**-supervised learning the "**self**", i.e. the model/data, is all you need
- Instead, careful data curation – *BY HUMANS* – is king 👑
  - ***<u>We</u>*** pick sentences (g.t. prefixes and labels) to maximize downstream perf
  - This process is nothing more than highly efficient batch labeling
- Ok, this point is mostly a semantic quibble